

ЛЕКЦІЯ 2. МЕТОД ВІДХИЛЕНЬ

НАВЧАЛЬНІ ПИТАННЯ

Попередні перетворення вихідних даних

Алгоритм методу відхилень

1. Попередні перетворення вихідних даних

Раніше, при розбитті даних на однорідні сукупності, потрібно було провести перетворення даних. В процедурі, що пропонується нижче, потрібно ввести додатковий розподіл ознак на стимулятори, дестимулятори та номінатори. В загальному випадку вони визначаються таким чином.

Ознака X називається стимулятором, якщо

$$\{x_i \geq x_j\} \Rightarrow \{n_i \succ n_j\}; \quad (2.1)$$

Тобто, якщо значення ознаки для одного об'єкта більше за значення цієї ознаки для іншого об'єкта випливає домінування першого об'єкта над іншим (більше значення ознаки відповідає "кращій" якісній характеристиці об'єкта). Аналогічно, для дестимулятора:

$$\{x_i \geq x_j\} \Rightarrow \{n_i \prec n_j\}; \quad (2.2)$$

і для номінатора

$$\begin{aligned} \{x_i \geq x_j \geq x_k\} &\Rightarrow n_i \succ n_j \\ \{x_k \geq x_i \geq x_j\} &\Rightarrow n_i \prec n_j, \end{aligned} \quad (2.3)$$

або

$$\begin{aligned} \{x_i \geq x_j \geq x_k\} &\Rightarrow n_i \prec n_j \\ \{x_k \geq x_i \geq x_j\} &\Rightarrow n_i \succ n_j \end{aligned}$$

тобто, спочатку ознака веде себе, як стимулятор (дестимулятор), а після досягнення деякого значення – як дестимулятор (стимулятор).

Вищезгадані типи ознак уведені для того, щоб врахувати їхній "економічний" зміст. Наприклад, ознака, що виражає продуктивність праці, є стимулятором, а ознака, яка виражає собівартість продукції – дестимулятором. Розглянемо ознаку, що описує якість виробленого виробу. При виробництві цигарок встановлений певний рівень вологості. Поки цей рівень не досягнутий, ознака має характер стимулятора; оптимальною точкою є граничний (плановий) рівень вологості. Далі ознака поводить себе як дестимулятор.

Для того щоб проводити подальший аналіз, потрібно перетворити дестимулятори в стимулятори. Ця операція виробляється відповідно до формули.

$$\begin{aligned} x'_i &= -x_i \\ x'_i &= \frac{1}{x_i} \end{aligned}, \quad (2.4)$$

Номінатори замінюються ознаками, що мають близьку змістовну інтерпретацію, але в той же час є стимуляторами або дестимуляторами.

Перетворення дестимуляторів у стимулятори необхідно тому, що в пропонованому далі методі використовується ізоморфічне перетворення ознак. Воно дає позитивні результати тільки тоді, коли досліджувана сукупність ознак складається зі стимуляторів. У цьому випадку спостерігається так звана позитивна кореляційна залежність: ріст значень однієї ознаки викликає ріст значень інших ознак. А тоді, виділяючи ізоморфічно однорідні підмножини, одержимо саме ті, що мають бажані з погляду економетричних

досліджень властивості. Тобто, будуть одержані підмножини точок-об'єктів із близькими структурами значень ознак. У протилежному ж випадку підмножина однорідних в смислі структури значень ознак виявиться розділеною, що небажано.

1. Алгоритм методу відхилень

Як відзначалось раніше, в основі методу лежить ізоморфічне перетворення даних. Ідея методу полягає в переході від багатомірних об'єктів до двовимірних. Це дає можливість графічно зобразити результат і провести попередній візуальний розподіл сукупності об'єктів на однорідні підмножини.

Алгоритм методу укладається в послідовність таких кроків.

1 крок. Перетворення дестимуляторів у стимулятори.

2 крок. Визначення полюсів сукупності. Полюсом є точка, яка має найбільші (верхній полюс) або всі найменші (нижній полюс) значення ознак. При цьому така точка взагалі може не входити в сукупність даних.

Позначимо верхній полюс через $P_B(x_1^*, x_2^*, \dots, x_n^*)$, а нижній через $P_H(x_1^*, x_2^*, \dots, x_n^*)$.

3 крок. Перенос системи координат у нижній полюс. При цьому значення вихідних даних перетворюється за формулою

$$u_{ij} = x_{ij} - x_i^* \quad (2.5)$$

Слід зауважити, що координати нижнього та верхнього полюсів також перетворюються відповідно до (2.5), однак для зручності позначення їх координат ми залишимо без змін

4 крок. Здійснення побудови осі сукупності даних. Вісь має таке рівняння

$$y_j = x_j^* \cdot t, \quad (2.6)$$

де t – параметр прямої. Ця пряма являє собою найбільшу вісь сукупності в багатомірному просторі. Відхилення від точок-спостережень від цієї осі дають інформацію про їх розташування.

5 крок. Знаходження прямокутних проекції точок на вісь. Позначимо їх через $R_i = (y_{i1}, y_{i2}, \dots, y_{in})$. Координати проекції обчислюються за формулою

$$y_{ij} = x_j^* \cdot t_i, \quad (2.7)$$

де

$$t_i = \frac{\sum_{j=1}^n x_j^* u_{ij}}{\sum_{j=1}^n (x_j^*)^2}. \quad (2.8)$$

6 крок. Визначення відхилень точок-об'єктів від осі сукупності. Будемо вважати, що лінійне упорядкування об'єктів P_i відповідає упорядкуванню проекцій R_i цих об'єктів. Отже, воно збігається з черговістю розташування точок R_i на осі сукупності, тобто з порядком, що відповідають відстаням точок від нижнього полюса P^* . В ролі відхилень виступають показники

$$\begin{aligned} M^* &= [(R_i - P_*) (R_i - P_*)^T]^{1/2} \\ W^* &= [(P_i - R_i) (P_i - R_i)^T]^{1/2}. \end{aligned} \quad (2.9)$$

Компоненти цих показників обчислюються за формулами

$$m_s^* = \sqrt{\sum_{j=1}^n y_{sj}^2};$$

$$w_s^* = \sqrt{\sum_{j=1}^n (x_{sj} - y_{sj})^2};$$
(2.10)

Перший показник задає відстань точок-проекцій від нижнього полюса (в новій системі – від початку координат), а другий – відстань між точками-об’єктами та їх прямокутними проекціями на вісь сукупності. Величини M^* W^* рекомендується нормувати, щоб їх значення лежали в межах від 0 до 1.

$$m_s = \frac{m_s^*}{\max_s \{m_s^*\}}; \quad w_s = \frac{w_s^*}{\max_s \{w_s^*\}}.$$
(2.11)

7 крок. Побудова в площині *MOW* діаграми розсіювання ознак. По її виду визначається слід розподілу ознак. Помітимо, що для багатомірних величин важко визначити знак (+ чи –) для відхилень точок сукупності. Тому Діаграма розсіювання зображається лише в першому квадранті системи координат *MOW*. Враховуючи цю особливість. Форма сліду розподілу буде визначена наближено.

Таким чином, від багатомірних точок-об’єктів шляхом перетворень здійснюється перехід до їх зображення на площині, що спрощує візуальний аналіз сліду розподілу.

Якщо слід еліпсоїдальний, то вся сукупність даних однорідна.

Якщо слід С-образний , то сукупність розбивається на дві однорідні підмножини.

Якщо слід S-образний, то він спочатку ділиться на два С-образних, а за їх видом проводиться розбиття сукупності даних.

Приклад. Завдання 2.1.

Нехай відомі такі дані по 10 підприємствам: X_1 – продуктивність праці, X_2 – індекс зниження собівартості продукції. Потрібно розбити сукупність даних на однорідні підмножини. Вихідні дані наведені в табл. 2.1.

Таблиця 2.1

Підприємство, n_i	Продуктивність праці, X_1	Індекс зниження собівартості продукції, X_2
1	5,88	52,6
2	6,3	46,6
3	8,17	172,1
4	9,12	56,5
5	9,26	204,2
6	9,35	62
7	9,38	209,6
8	9,87	53,1
9	10,81	236,7
10	12,11	222,6

Розв’язок.

Виконаємо розрахунки відповідно до алгоритму, наведеного в описі методу. Вихідні дані є стимуляторами, тому крок 1 пропускаємо . Результати виконання кроків 2,3 та 4 занесемо до табл. 2.2. Діаграма розсіювання точок-об’єктів зображена на рис 2.1.

	35,96	0,00	15752,17	100,02	24846,00	240,42	26577,86	43,92	36160,63	31014,60	
M_s^*	6,00	0,01	125,51	10,00	157,63	15,51	163,03	6,63	190,16	176,11	max= 190,16
M_s	0,03	0,00	0,66	0,05	0,83	0,08	0,86	0,03	1,00	0,93	
$(U_1 - Y_1)^2$	0,04	0,18	3,32	8,48	3,18	8,77	3,39	14,23	1,69	0,21	
$(U_2 - Y_2)^2$	0,00	0,00	0,00	0,01	0,00	0,01	0,00	0,02	0,00	0,00	
W_s^*	0,20	0,42	1,82	2,91	1,78	2,96	1,84	3,77	1,30	0,46	max= 3,77
W_s	0,05	0,11	0,48	0,77	0,47	0,79	0,49	1,00	0,34	0,12	

Зобразимо точки-об'єкти в системі координат MOW (рис 2.2). В результаті за виглядом одержаного зображення можна зробити такий висновок про однорідність даних.

Вся сукупність розбивається на 4 ізоморфічно однорідних підмножини:

$$\Omega_1 = \{ n_1, n_2 \}.$$

$$\Omega_2 = \{ n_3, n_5, n_7 \}.$$

$$\Omega_3 = \{ n_4, n_6, n_8 \}.$$

$$\Omega_4 = \{ n_9, n_{10} \}.$$

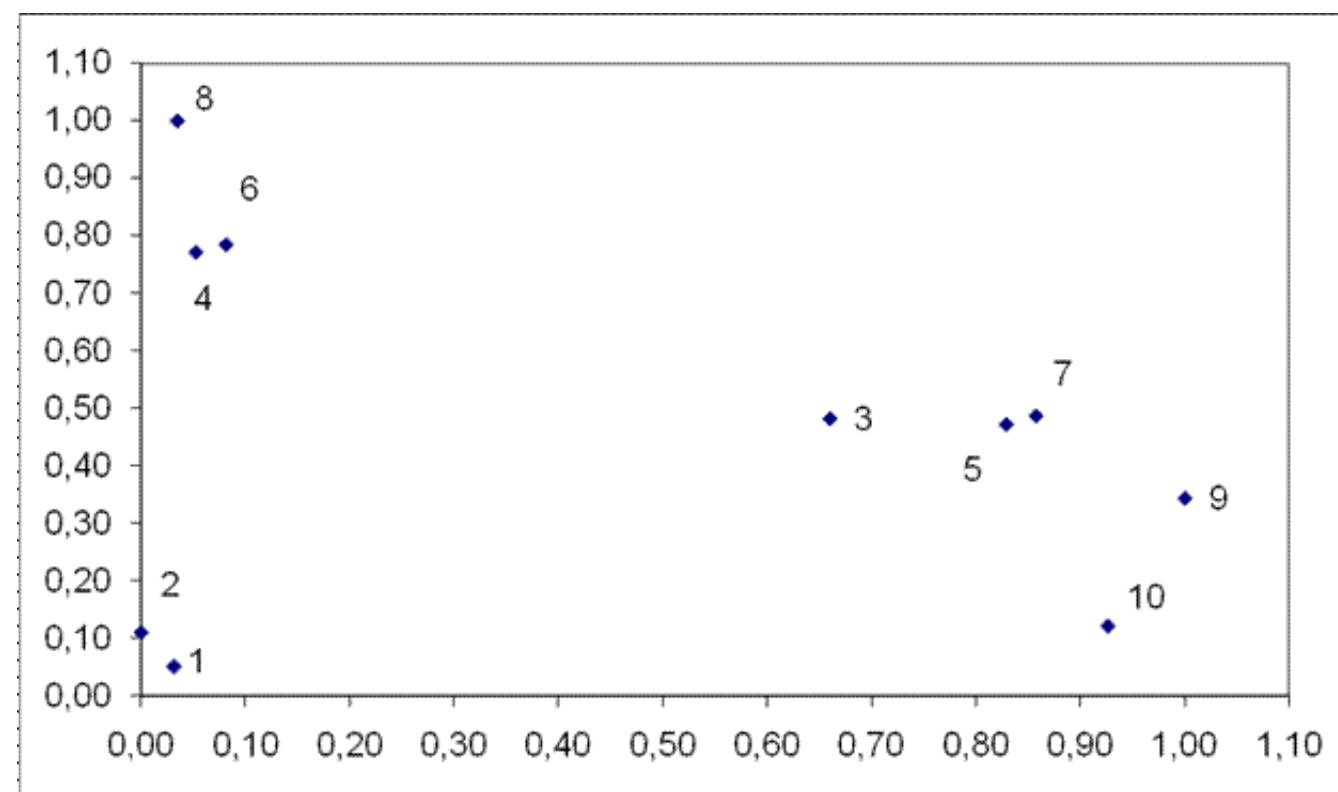


Рис 2.2. Зображення об'єктів в системі координат MOW

2. Питання для самоперевірки.

1. Що собою являють стимулятори, дестимулятори та номінатори?
2. В чому сутність методу відхилень?
3. Охарактеризуйте основні кроки методу відхилень.